

Cuestionarios conceptuales en Física: evaluación de resultados de aprendizaje y conceptos erróneos en Mecánica y Electromagnetismo

Concept Inventories in Physics: assessment of learning outcomes and misconceptions in Mechanics and Electromagnetism

F. Alconchel, M. E. Cámara, M. Díaz, B. Gámez, L. Gámez, M. F. Laguna, Á. Lavín, P. Martín, Á. Ponce, L. Seidel
email luis.seidel@upm.es

Departamento de Física Aplicada e Ingeniería de Materiales
E.T.S.I. Industriales. Universidad Politécnica de Madrid
Madrid, España

Resumen- En este trabajo presentamos los resultados de la aplicación de dos cuestionarios conceptuales (FCI, Force Concept Inventory, y BEMA, Brief Electricity and Magnetism Assessment) a estudiantes de primer curso de Grados en Ingeniería en la Universidad Politécnica de Madrid. Hemos utilizado los cuestionarios conceptuales como una herramienta de investigación educativa orientada a identificar fortalezas y debilidades en el perfil de entrada de los alumnos y en los resultados de aprendizaje de las asignaturas de Física General I y II. Además, hemos identificado los errores conceptuales más importantes. Los resultados obtenidos nos permiten dirigir las acciones de innovación educativa a objetivos mucho más concretos. Hemos comprobado que los cuestionarios conceptuales son una herramienta válida para complementar la evaluación de los resultados de aprendizaje de la Física en Ingeniería.

Palabras clave: cuestionarios conceptuales, física, evaluación de resultados de aprendizaje, conceptos erróneos

Abstract- In this work we present the results of the administration of two well-known Concept Inventories (FCI, Force Concept Inventory and BEMA, Brief Electricity and Magnetism Assessment) to first-year students in Engineering Degrees at Universidad Politécnica de Madrid. We have used the concept inventories as a tool for education research aimed at identifying strengths and weaknesses in the input profile of students and in the learning outcomes of the subjects General Physics I and II. In addition, we have identified the most important misconceptions. The results allow us to direct the actions of educational innovation to more specific targets. We have found that concept inventories are a valid tool to complement the assessment of learning outcomes of Physics in Engineering.

Keywords: concept inventories, physics, assessment of learning outcomes, misconceptions

1. INTRODUCCIÓN

La innovación educativa necesita apoyarse sobre la investigación educativa. Para mejorar el proceso de enseñanza-aprendizaje, hay que conocer cómo está funcionando en un determinado momento ese proceso. En los distintos campos del conocimiento se han ido desarrollando herramientas para evaluar y diagnosticar la transmisión de conocimiento del profesor al estudiante, o el desarrollo del aprendizaje en el propio estudiante. En el campo concreto de

la Física, desde hace unos 30 años se ha consolidado una línea de investigación que se conoce como Physics Education Research (PER) en la que, con métodos tomados tanto de la Física como de la Educación, y con foros comparables a la de cualquier rama científica (revistas científicas, congresos,...) se está progresando en el estudio experimental de cómo se aprende Física en las distintas etapas educativas, en particular en la Universidad. Se puede ver un repositorio completo de actividades, patrocinado por la AAPT (asociación americana de profesores de física) en www.compadre.org/per/.

Una de las herramientas más utilizadas en PER son los cuestionarios conceptuales (Concept Inventories - CI), baterías de preguntas de elección múltiple desarrolladas y validadas experimentalmente con una metodología que en parte proviene de la psicometría. El motivo principal de la popularidad de los CI es el éxito que ha tenido el primero de los desarrollados en el ámbito de la Física, el Force Concept Inventory (FCI) (Hestenes, Wells, & Swackhamer, 1992). Este cuestionario consta, en la versión actual disponible desde 1995, de 30 preguntas de elección múltiple, cada una de ellas con 5 opciones, solo una de ellas correcta. Los conceptos que trata de evaluar son los que se refieren a la Mecánica newtoniana clásica, en particular el concepto de fuerza. Durante más de 20 años, el FCI se ha utilizado en casi todas las universidades norteamericanas y en muchas de todo el mundo (Hake, 1998). Una revisión de resultados en universidades españolas se puede ver en (Covián y Celemín, 2008). Un estudio orientado a la identificación de errores conceptuales dominantes es (Martín-Blas, Seidel, & Serrano-Fernández, 2010).

Desde la aparición del FCI se han desarrollado CI en casi todas las áreas de la Física y las ciencias. Algunos ejemplos son: TCCI para Termodinámica, ECA para Energía, ASSCI en Astronomía, TUV para vectores (Barniol & Zavala, 2014) o RCI para relatividad (Aslanides & Savage, 2013).

Existen varios CI para Electromagnetismo, de distinto nivel. Para contenidos de primer curso de Grado los más conocidos son BEMA (Brief Electricity and Magnetism Assessment) desarrollado en 1997 (Ding, Chabay, Sherwood, & Beichner, 2006) y CSEM (Conceptual Survey of Electricity and Magnetism) desarrollado a partir de 1995 (Maloney, O'Kuma,

Hieggelke, & Van Heuvelen, 2001). Ambos han sido utilizados en estudios con gran número de alumnos. Hemos escogido el BEMA porque existe una versión traducida al español.

La validación de un CI utiliza técnicas estadísticas centradas en dos conceptos: confiabilidad del test y poder discriminatorio. Ambos se refieren tanto a las preguntas individuales como al cuestionario en conjunto. Un cuestionario es confiable si arroja resultados similares cuando se aplica a poblaciones similares en las mismas condiciones. Un cuestionario tiene buen poder discriminatorio cuando sus resultados permiten distinguir claramente a los estudiantes que tienen un buen conocimiento del ámbito conceptual del CI de los que no lo tienen. Todos los CI mencionados, y mucho más el FCI, tienen más que comprobadas ambas propiedades. Sin embargo, nunca es superfluo comprobar en una nueva administración del FCI que las propiedades estadísticas que definen esas propiedades siguen estando en el rango adecuado, antes de sacar otras conclusiones de los resultados. (Ding & Beichner, 2009). Los CI también han recibido críticas, en cuanto al valor que se le puede dar a los resultados globales del test y al análisis de los mismos basados exclusivamente en la teoría clásica de los CI (Wallace & Bailey, 2010).

El BEMA evalúa conceptos clásicos del electromagnetismo: electrostática, circuitos de corriente continua, magnetostática, inducción electromagnética. Consta de 31 preguntas de elección múltiple, pero a diferencia del FCI el número de opciones es variable, entre 3 y 10. Algunas preguntas que se refieren a la misma situación física se corrigen conjuntamente.

La validación del BEMA se ha llevado a cabo en estudios como (Ding, 2014; Kohlmyer et al., 2009; Pollock, 2009). También existen trabajos que analizan la presencia de errores conceptuales en Electromagnetismo, pero utilizando otro CI (CSEM) (Leppävirta, 2012).

2. CONTEXTO

La investigación que se presenta en este trabajo se engloba en un proyecto de innovación educativa cuyo objetivo principal es definir los resultados de aprendizaje de las asignaturas de Física General de primer curso de los Grados en Ingeniería que se imparten en la ETSI Industriales de la Universidad Politécnica de Madrid y evaluar su adquisición con los métodos de evaluación que se están utilizando. Este objetivo es un primer paso para detectar carencias que la evaluación no detecta y para introducir metodologías innovadoras que remedien tales carencias y optimicen la eficacia del proceso de enseñanza-aprendizaje.

La metodología empleada hasta ahora en el aula (lección magistral, enfoque expositivo basado en desarrollos teóricos, ejemplos y problemas excesivamente matemáticos y operacionales) y los métodos de evaluación (examen final con énfasis en la resolución de ejercicios apoyados en fórmulas y leyes físicas, pocas pruebas de evaluación continua con un formato similar) se podrían denominar tradicionales. Esta metodología tiene sus fortalezas y sus ventajas, sobre todo en grupos de clase muy numerosos, pero es bien conocido que tiene también sus lagunas. En particular, hay cada vez más estudiantes que son capaces de resolver un ejercicio que exija

memorizar fórmulas y realizar cálculos de mediana dificultad, sin haber entendido nada de la situación física que plantea el ejercicio; por tanto, pueden superar la asignatura, sin haber adquirido los resultados de aprendizaje de los primeros niveles del plano cognitivo de la taxonomía de Bloom (nivel de comprensión, en particular) (Kennedy, 2007).

Ante este problema, la utilización de los CI como herramienta para comprobar experimentalmente esa sospecha se nos hizo evidente. Nos planteamos entonces hacer una investigación basada en la aplicación del FCI en la asignatura de Física General I, cuyos resultados de aprendizaje están centrados en la Mecánica newtoniana clásica. Generalmente, los CI se administran al comienzo y al final de un periodo de aprendizaje para evaluar el progreso en ese tiempo. Para complementar la investigación basada en el FCI para la asignatura de Física General I, decidimos llevar a cabo otra experiencia similar basada en el BEMA (Brief Electricity and Magnetism Assessment) para la asignatura de Física General II.

La aportación original de nuestro trabajo está en la aplicación conjunta de una serie de indicadores estadísticos para identificar de forma cuantitativa los modelos mentales que subyacen a las respuestas y los errores conceptuales dominantes, siguiendo la metodología de (Martín-Blas, Seidel, & Serrano-Fernández, 2010).

3. DESCRIPCIÓN

En el curso 2014/15 se ha administrado el FCI en el primer semestre y el BEMA en el segundo semestre a los alumnos de primer curso de los Grados en Ingeniería en Tecnologías Industriales (GITI) e Ingeniería Química (GIQ) de la ETSI Industriales de la Universidad Politécnica de Madrid. Hay seis grupos de clase de GITI y uno de GIQ.

Como es habitual, los cuestionarios se administran dos veces: una al comienzo del semestre (septiembre o febrero, PRE) y otra al finalizar (diciembre o mayo, POST). Los alumnos del GITI han accedido a la Universidad con una nota mínima de 11.7 (sobre 14) y los del GIQ de 11.015. El perfil de entrada de los alumnos destaca por la elevada nota de corte, que hace que en conjunto sean alumnos excelentes, con elevada motivación y capacidad de trabajo. En general, también valoran la Física como muy importante en su formación. La diferencia más significativa entre los alumnos de GITI y GIQ es que un porcentaje significativo (en torno al 20%) de estos últimos no han cursado física en bachillerato. Los contenidos de las asignaturas de Física General I y II son muy similares para ambos Grados y cubren los contenidos estándar de Física Universitaria.

El cuestionario se realiza de forma voluntaria y anónima, en el aula de clase, durante el tiempo recomendado (30 ó 35 minutos). Se pide a los estudiantes que respondan a todas las cuestiones, sin dejar ninguna en blanco. También se les solicita que pongan un identificador (su número de matrícula) en la hoja de respuestas, con el fin de poder analizar su evolución personalizada entre los dos tests. Los profesores responsables de la administración realizaron un esfuerzo adicional para motivar a los estudiantes a que se tomaran en serio el cuestionario. El resultado de los tests no ha tenido ninguna influencia sobre la evaluación de la asignatura.

No se han tenido en cuenta en los análisis posteriores los cuestionarios con más de 10 ítems sin contestar. No se ha recogido información sobre sexo o edad de los estudiantes, y se pidió que completaran el cuestionario solo los alumnos de nuevo ingreso. Se trató de minimizar las diferencias entre grupos y las condiciones ambientales, que se sabe que pueden aparecer y tener efectos sobre los resultados (Ding, Reay, Lee, & Bao, 2008).

4. RESULTADOS

A. Force Concept Inventory

El número de alumnos que han realizado el cuestionario FCI ha sido de 371 (GITI) y 67 (GIQ) en septiembre y 322 (GITI) y 44 (GIQ) en diciembre. En la Tabla 1 se recogen los resultados globales: nota media, desviación típica e índice de dificultad.

Tabla 1. Resultados globales del FCI.

	Nº alumnos	Nota media	Desviación típica	Índice dificultad
GITI PRE	371	52,40%	16,93%	0,546
GIQ PRE	67	34,38%	14,81%	0,352
GITI POST	322	61,25%	14,93%	0,625
GIQ POST	44	53,18%	20,47%	0,540

En la Figura 1 se muestra la distribución de notas global en septiembre (PRE) y diciembre (POST).

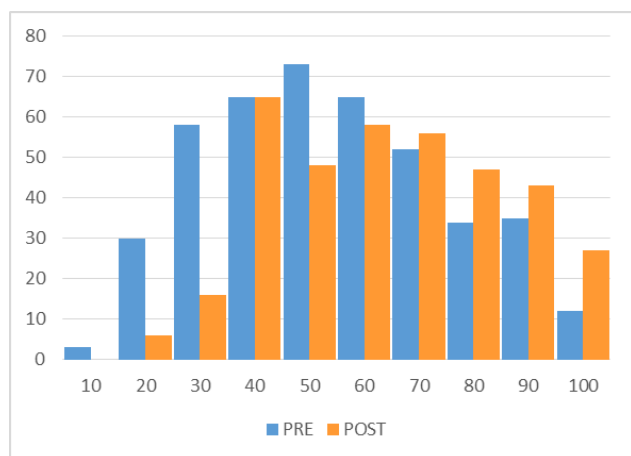


Figura 1. Distribución de notas en el FCI, número de alumnos en cada decena de porcentaje total. Los resultados del POST aparecen desplazados hacia la derecha, indicando la mejora global de resultados.

Los resultados se obtuvieron separados por grupos de clase. Un análisis posterior de las diferencias entre grupos (prueba t de Student) confirma que las diferencias no son estadísticamente significativas entre grupos del GITI pero sí lo son entre GITI y GIQ, por lo que ambos resultados se presentan separados.

Se conoce como índice de dificultad de un ítem de un cuestionario conceptual al cociente entre el número de estudiantes que responden correctamente a un ítem y el número total de estudiantes que realizan el cuestionario. Por tanto, el índice de dificultad es un número entre 0 (ningún acierto) y 1 (100% de aciertos). Hay que tener en cuenta que preguntas más fáciles tienen un índice de dificultad mayor. Hemos mantenido la definición que se encuentra en toda la literatura para facilitar las comparaciones. El índice de dificultad del CI es la media de los valores de los ítems. El índice de dificultad obtenido y reflejado en la Tabla 1 está en el rango que se considera óptimo, entre 0,3 y 0,8. Adicionalmente, el índice de dificultad de todas las preguntas está en el mismo rango, en el POST. Estos resultados confirman la confiabilidad del FCI.

El índice de discriminación se utiliza para comprobar si un ítem del CI discrimina entre los alumnos con mejores resultados y los que tienen peores resultados. La definición más utilizada es: $D_{25} = (N_H - N_L) / (N/4)$, donde N_H es el número de alumnos con nota global en el cuartil superior (25% mejor) que aciertan el ítem, N_L lo mismo para los que están en el cuartil inferior (25% peor), y N el número total de alumnos. Su valor está comprendido entre -1 y 1. Se consideran satisfactorios valores por encima de 0,3. El índice de discriminación del CI es la media de los índices de los ítems. En nuestro estudio, todas las cuestiones obtuvieron un índice de discriminación (en el POST) positivo, y todas salvo una por encima de 0,3.

La ganancia absoluta es la diferencia entre las notas porcentuales del POST y PRE. Una medida más útil es la ganancia relativa, g , definida por (Hake, 1998) que es la ganancia absoluta dividida por la máxima ganancia posible, $100 - \text{PRE}$. El valor global de g se ha utilizado para clasificar los cursos y métodos utilizados en el aprendizaje. En nuestro estudio, la ganancia relativa total fue de 0,186 (GITI) y 0,286 (GIQ). Estos valores se comparan bien con otros publicados anteriormente. En cursos tradicionales, los valores de g están en el rango 0,2-0,3.

En el análisis de cuestiones individuales se ha calculado la concentración definida por (Bao & Redish, 2001), tanto para todas las respuestas, como para las incorrectas. Un valor próximo a 1 de la concentración indica un ítem en el que los alumnos se inclinan mayoritariamente por una respuesta (correcta o no). Un valor pequeño, próximo a 0, corresponde a una dispersión en las respuestas, que sugiere que probablemente la cuestión ha sido respondida al azar. Un valor próximo a 1 en la concentración de respuestas incorrectas indica la muy probable aparición de un concepto erróneo. Estudiar la evolución de la concentración del PRE al POST permite identificar las cuestiones que presentan conceptos erróneos persistentes. En nuestro estudio, hemos comprobado la presencia de los siguientes conceptos erróneos, según la clasificación de (Martín-Blas et al., 2010):

- Fuerza paralela a la velocidad (Q13c, Q17a, Q25d)
- Pares acción/reacción no iguales (Q4a, Q15c)
- No distinguir entre fuerza e impulso (Q30e)
- Tiempo de desplazamiento depende de la masa (Q2d)

Por último, hemos estudiado la correlación entre los resultados del FCI POST y la nota del examen final, para los

210 alumnos del GITI que se identificaron al realizar los cuestionarios. Se ha obtenido un coeficiente de correlación de 0,344.

B. Brief Electricity and Magnetism Assessment

El número de alumnos que han realizado el cuestionario BEMA ha sido de 338 (GITI) y 55 (GIQ) en febrero y 242 (GITI) y 43 (GIQ) en mayo. En la Tabla 2 se recogen los resultados globales: nota media, desviación típica e índice de dificultad.

Tabla 2. Resultados globales del BEMA.

	Nº alumnos	Nota media	Desviación típica	Índice dificultad
GITI PRE	338	34,29%	19,27%	0,337
GIQ PRE	55	28,67%	20,08%	0,276
GITI POST	242	44,70%	23,64%	0,437
GIQ POST	43	40,00%	23,62%	0,393

En la Figura 2 se muestra la distribución de notas global en febrero (PRE) y mayo (POST).

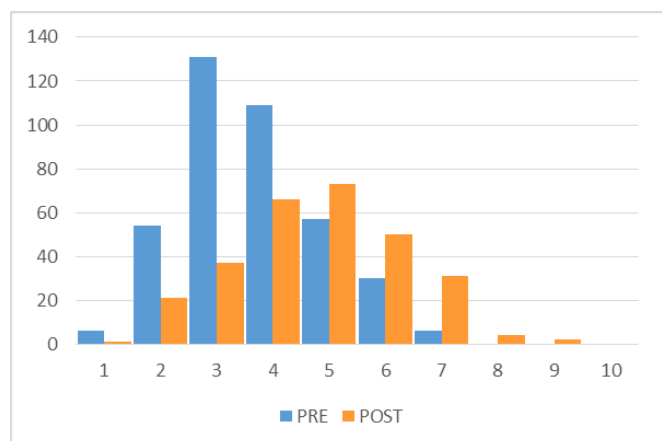


Figura 2. Distribución de notas en el BEMA, de 0 a 10 y número de alumnos en cada decena de porcentaje total. Los resultados del POST aparecen desplazados hacia la derecha, indicando la mejora global de resultados.

El BEMA es un CI considerablemente más difícil que el FCI, lo que se pone de manifiesto en los resultados, tanto PRE como POST. En la distribución por notas se observa que muy pocos alumnos obtienen notas altas, incluso en el POST. Esto se debe por un lado a la extensión del test y por otro a que contiene preguntas sobre conceptos que no se cubren en la asignatura de Física General II, como electrolisis o algunas cuestiones de circuitos de corriente continua.

El índice de dificultad está cerca del límite de validez para un CI (en torno al 0,3-0,4) y además hasta 9 ítems (en el POST) tienen un índice de dificultad por debajo de 0,2, lo que los convierte en preguntas muy difíciles, que se han respondido al azar y cuyo análisis no nos va a permitir

encontrar ninguna información adicional. Se trata de las cuestiones relacionadas con los temas antes mencionados.

Para el índice de discriminación, resulta negativo para dos cuestiones, 28 y 29, relacionadas con la inducción electromagnética. Como se ha mencionado antes, tales cuestiones tienen que descartarse de un análisis más detallado. Para el resto de cuestiones, el índice de discriminación es positivo (lo que significa que esas cuestiones discriminan adecuadamente) pero solo es mayor de 0,5 para 6 cuestiones.

La ganancia relativa de Hake es de 0,158 tanto para GITI como para GIQ. El valor tan bajo obtenido puede deberse tanto a la dificultad del CI y al número de cuestiones que hay que descartar como a la mayor presencia de errores conceptuales, que se analizan a continuación. Sin embargo, la ganancia relativa de las cuestiones individuales es significativamente alta (por encima de 0,4) para las cuestiones de electrostática y magnetostática, que son las más ligadas a los contenidos de Física General II. Se encuentran algunas cuestiones de ganancia negativa (es decir, con resultados peores en el POST que en el PRE), lo que debe explicarse por ser cuestiones contestadas al azar.

La concentración, tanto global como de respuestas incorrectas, debe interpretarse como una forma de clasificar las cuestiones por la dispersión de opciones. Una cuestión con una concentración próxima a 1, en la respuesta correcta, es una cuestión clara y fácil que no necesita mayor análisis. Una concentración próxima a 0, con todas las opciones elegidas por un porcentaje parecido de alumnos, aparece en cuestiones contestadas al azar. Del primer tipo son las cuestiones 1 a 5 (electrostática), 20 y 21 (magnetostática). Hay que notar que en el caso del BEMA el análisis se complica porque muchas cuestiones tienen un gran número de opciones (entre 7 y 10) y porque realmente no tiene distractores como en el caso del FCI (opciones incorrectas elegidas para poner de manifiesto un concepto erróneo) sino que simplemente presenta como opciones todas las posibilidades que se pueden dar (por ejemplo, la dirección y sentido de una fuerza, o un campo en el espacio tiene al menos 6 posibles valores perpendiculares entre sí). Sin embargo, la concentración calculada como se define en (Bao & Redish, 2001) da un valor independiente del número de opciones del ítem.

Por último, el análisis de la concentración de respuestas incorrectas nos permite identificar errores conceptuales evidentes en las siguientes cuestiones:

- Q12d (campo eléctrico nulo en el interior de un conductor debido a la corriente)
- Q14a (signo incorrecto de la diferencia de potencial)
- Q17a (diferencia de potencial nula entre los bornes de una batería)
- Q23f (sentido incorrecto del campo)

Varios de estos errores no son realmente conceptuales, sino errores en el signo o sentido de una magnitud. Están asociados al uso incorrecto del producto vectorial o del signo de una carga eléctrica (por ejemplo, la carga de un electrón es negativa).

Hemos estudiado también la correlación entre los resultados del BEMA y el examen final de la asignatura Física General II. Como ha ocurrido en el caso del FCI, la correlación entre la

nota del BEMA y del examen final no es muy alta. En este caso, el coeficiente de correlación es 0,338.

Se ha realizado también una comparación entre los resultados del FCI y el BEMA. Para los 142 alumnos identificados que han realizado los 4 tests, 112 alumnos tienen ganancia positiva en ambos, 19 alumnos tienen nota mayor al 50% en todos los tests y 29% tienen nota inferior al 50% en todos los tests. La ganancia media en ambos tests es 0.1. Sin embargo, la muestra es reducida y sesgada respecto a la población total, por lo que no se pueden extraer más conclusiones.

5. CONCLUSIONES

Los resultados obtenidos en esta investigación utilizando el FCI corroboran la utilidad de los CI. Los resultados globales están en el rango esperado tanto para los valores medios como para ganancia, índice de dificultad y de discriminación. Hemos verificado la solidez y confiabilidad del FCI. Pero más interesantes son los resultados obtenidos para la concentración y la presencia de conceptos erróneos. En este caso, que frecuentemente no se analiza en detalle, se ponen de manifiesto unos pocos conceptos erróneos, bien identificados. En esta situación es más fácil diseñar metodologías o herramientas que se dirijan directamente a estos conceptos. La poca eficacia del aprendizaje en la asignatura de Física General I para corregir esos conceptos erróneos, que son persistentes, muestra que están asociados a un modelo mental de la mecánica que podríamos llamar pre-newtoniano. Nuestra experiencia es que es útil seguir aplicando el FCI para comprobar la evolución de estos conceptos erróneos y su universalidad, es decir, su aparición en todos los alumnos de la misma formación previa y el mismo nivel.

AGRADECIMIENTOS

Este trabajo se ha desarrollado en el Proyecto de Innovación Educativa IE1415-05019 "LOP - Learning Outcomes in Physics – Resultados de Aprendizaje en Física" de la Universidad Politécnica de Madrid. Agradecemos el trabajo de Pablo Álvarez Blanco en el tratamiento de los datos.

REFERENCIAS

Aslanides, J. S., & Savage, C. M. (2013). Relativity concept inventory: Development, analysis, and results. *Physical Review Special Topics - Physics Education Research*, 9(1), 010118.

Bao, L., & Redish, E. F. (2001). Concentration analysis: A quantitative assessment of student states. *American Journal of Physics*, 69(S1), S45.

Barniol, P., & Zavala, G. (2014). Test of understanding of vectors: A reliable multiple-choice vector concept test.

Physical Review Special Topics - Physics Education Research, 10(1), 010121.

Ding, L. (2014). Seeking missing pieces in science concept assessments: Reevaluating the Brief Electricity and Magnetism Assessment through Rasch analysis. *Physical Review Special Topics - Physics Education Research*, 10(1), 010105.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics - Physics Education Research*, 2(1), 010105.

Ding, L., Reay, N. W., Lee, A., & Bao, L. (2008). Effects of testing conditions on conceptual survey results. *Physical Review Special Topics - Physics Education Research*, 4(1), 010112.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141.

Kohlmyer, M., Caballero, M., Catrambone, R., Chabay, R., Ding, L., Haugan, M., ... Schatz, M. (2009). Tale of two curricula: The performance of 2000 students in introductory electromagnetism. *Physical Review Special Topics - Physics Education Research*, 5(2), 020105.

Leppävirta, J. (2012). The Effect of Naïve Ideas on Students' Reasoning About Electricity and Magnetism. *Research in Science Education*, 42(4), 753–767.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., & Van Heuvelen, A. (2001). Surveying students' conceptual knowledge of electricity and magnetism. *American Journal of Physics*, 69(S1), S12.

Martín-Blas, T., Seidel, L., & Serrano-Fernández, A. (2010). Enhancing Force Concept Inventory diagnostics to identify dominant misconceptions in first-year engineering physics. *European Journal of Engineering Education*, 35(6), 597–606.

Pollock, S. (2009). Longitudinal study of student conceptual understanding in electricity and magnetism. *Physical Review Special Topics - Physics Education Research*, 5(2), 020110.

Wallace, C. S., & Bailey, J. M. (2010). Do Concept Inventories Actually Measure Anything? *Astronomy Education Review*, 9(1), 010116.